

AN APPROACH TO CONSTRUCT MDRT TO PRODUCE QUALITATIVE RATINGS FOR E-COMMERCE WEBSITES

D. R. KUMAR RAJA¹ & S. PUSHPA²

¹Research Scholar, Department of CSE, St Peter's University, Avadi, Chennai, Tamil Nadu, India

²Professor & Head, Department of CSE, St Peter's University, Avadi, Chennai, Tamil Nadu, India

ABSTRACT

Online surveys, customer reviews on shopping sites are the key sources to understand customer requirements and feedback to help upgrade the product quality and achieve greater outcomes. In our previous paper, we targeted a novel approach to extract the customer sentiments or opinions at considerably much better granular level. However, there are many challenges in dealing with human languages. We only concentrated on reviews in English and the reviews which are much straight forward. In real world, customers display their emotions like anger, and try to be sarcastic sometimes. In addition to these challenges, we need to deal with different words to expressing same context. In this paper we want to take it forward to accept reviews from other languages and also address the problem of unknown words by making our system more adaptive.

KEYWORDS: Reviews, Ratings, Superior and Inferior Words, Parsing, MDRT

Received: Nov 15, 2016; **Accepted:** Jan 04, 2017; **Published:** Jan 07, 2017; **Paper Id.:** IJCSSEITRFEB20172

INTRODUCTION

Just like in the previous paper, we use a dictionary to get the meanings of the words in English. However, the current paper talks about, in addition to using the dictionary, updating the dictionary, getting the meaning of an unknown word to the system, understanding any idioms or predicting sarcasm in the comments and on top of all predicting the language in use and translating the reviews into English before analysis.

For getting the reviews of any product, we use Amazon APIs, flip kart APIs and also twitter APIs configured to flume (a tool used along with Hadoop). Once the reviews are flown into our HDFS, while processing, language of the reviews is identified then converted into English and then applied the algorithm for extracting the user opinion.

As per the statistical analysis ratings are considered based on structured and formal reviews. In this concern most of the reviewers are not considering un-formal and un-structured reviews due to the lack of regional languages and improper specifications of reviews like Emojie's, stickers, etc. While considering this kind of reviews for ratings, users are not in a position to judge whether the product is qualitative or not. And also some of the stack holders are getting difficulties for justifying the quality attributes of their products.

PROPOSED APPROACH

By considering all the constraints in reviews generation of the products we are going to propose a new novel approach that deals with both formal, un-formal and structured, un-structured reviews we can get qualitative

rating for the products in all the aspects. In this paper we concentrate mainly on un-formal reviews in the form of regional languages.

STRUCTURED APPROACH

For the conversion of regional languages we can use Natural Language Tool Kit in Python (NLTK). Natural Language Processing in Python provides a practical introduction to programming for language processing. Following lines of code briefs the actual processing.

Import Nltk

```
sentence = """At seven o'clock on Monday morning
Arthur didn't feel good."""
tokens = nltk.word_tokenize(sentence)

tokens

['At', 'seven', 'o'clock', 'on', 'Monday', 'morning',
'Arthur', 'did', 'n't', 'feel', 'good', '.']

tagged = nltk.pos_tag(tokens)

tagged[0:6]

[('At', 'IN'), ('seven', 'CD'), ('o'clock', 'JJ'), ('on', 'IN'),
('Monday', 'NNP'), ('morning', 'NN')]
```

Identify Named Entities

```
entities = nltk.chunk.ne_chunk(tagged)

entities

Tree('S', [(['At', 'IN'], ('seven', 'CD'), ('o'clock', 'JJ'),
('on', 'IN'), ('Monday', 'NNP'), ('morning', 'NN'),
Tree('PERSON', [(['Arthur', 'NNP'])],
('did', 'VBD'), ('n't', 'RB'), ('feel', 'VB'),
('good', 'JJ'), ('.', '.')])])

Display a parse tree:

from nltk.corpus import treebank

t = treebank.parsed_sents('wsj_0001.mrg')[0]

t.draw()
```

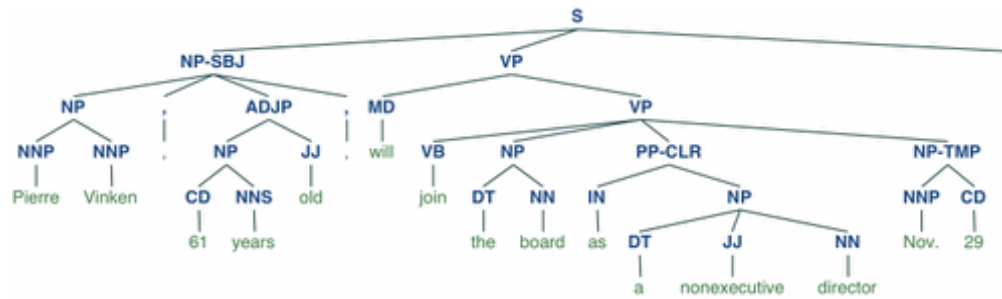


Figure 1

ANUSAARAKA

Anusaaraka is an English-Hindi language accessing software. It is a machine translation tool with insights from Panini's Ashtadhyayi (Method for Grammar rules); this will convert the Sanskrit language to English language.

Anusaaraka derives its name from the Sanskrit word 'Anusaran' which means 'to heed'. With the aim to reduce language barriers in the sentences Anusaaraka allows the user to import text in a language that is not known to the user. The present version of Anusaaraka provides for translation from English to Hindi.

Once you go to the of Anusaraka site you will be able to know more about grammar conversion and also be able give your suggestions to the Language Resource Development of this software, if you have comfortable knowledge of English and Hindi.

After converting the un-formal reviews into formal reviews the next phase of our work starts i.e mining the data using any one of the best practice mining tools

DEPENDENCY PARSING

Dependency Parsing is a technique which is used to identify the key terms in a review or in a paragraph. The identification process of key terms identification will be evaluated in three steps for dependency parsing

- Syntactic Representation
- Parsing Algorithm
- Machine Learning

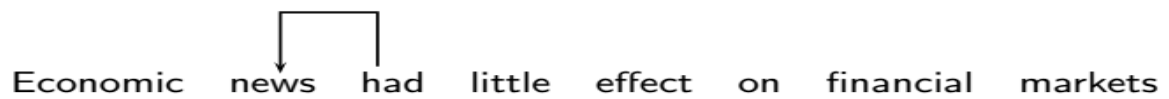
Syntactic Representation

Syntactic structure consists of lexical items linked by binary asymmetric relations called dependencies. In syntactic representation each sentence is organized as whole which belongs to elements of each word that belongs to a sentence case by itself to be isolated as in the dictionary.

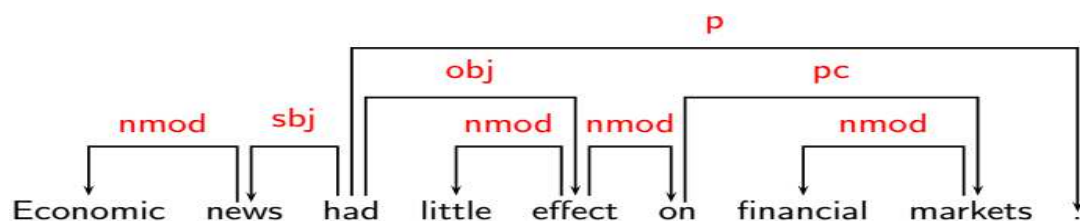
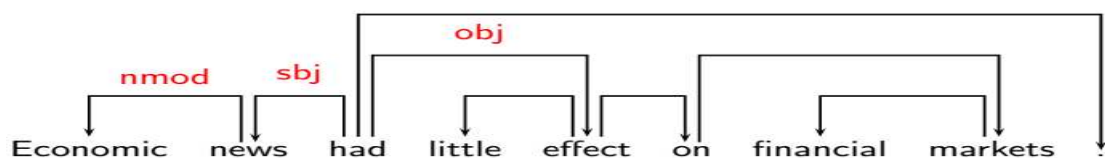
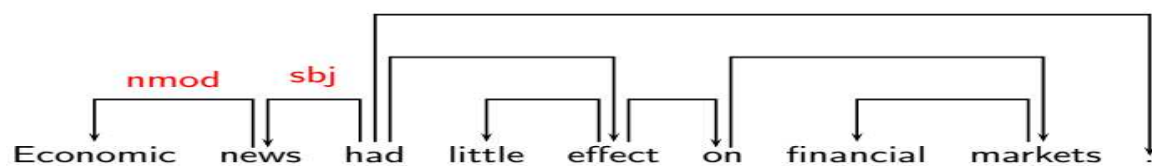
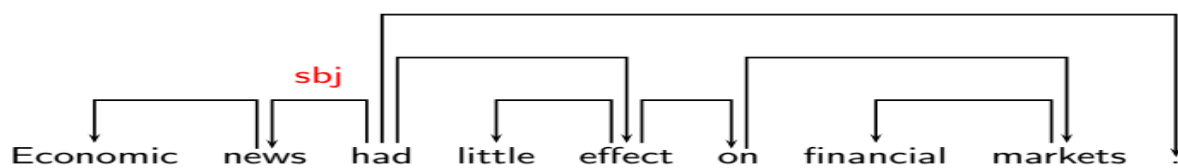
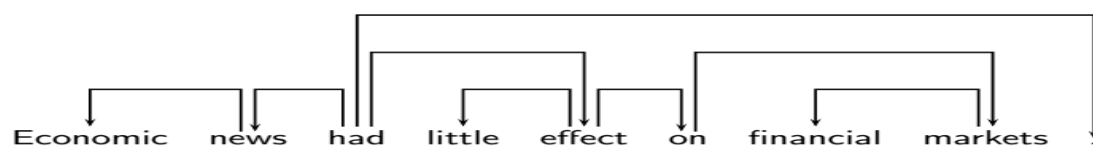
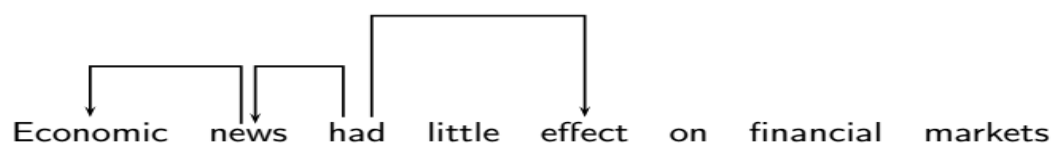
The structural connections established dependency relation between the words. Each connection in principle units consists a superior term and an inferior term.

Example for Syntactic Representation

Economic news had little effect on financial markets



From the above sentence “news” is a Noun will be act as superior term which is followed by “had” will be act as inferior term



Phrase Structure

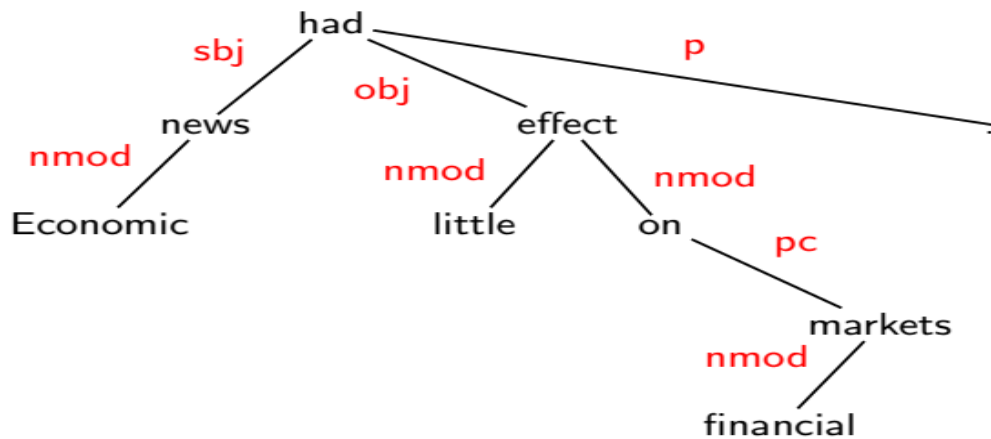


Figure 2

Parsing Algorithm

Training data: $\mathcal{T} = (sent_t, deps_t)_{t=1}^T$
 $\mathbf{w} = 0; \mathbf{v} = 0; i = 0;$
 for $n : 1..N$
 for $t : 1..T$
 $\mathbf{w}^{(i+1)} = \text{update } \mathbf{w}^{(i)} \text{ according to } (sent_t, deps_t)$
 $\mathbf{v} = \mathbf{v} + \mathbf{w}^{(i+1)}$
 $i = i + 1$
 $\mathbf{w} = \mathbf{v} / (N \cdot T)$

Statistical Evaluation of Parsing algorithm:

Table 1

Parser	English		Czech	
	W	S	W	S
<i>k</i> -best MIRA Eisner	90.9	37.5	83.3	31.3
best MIRA CLE	90.2	33.2	84.1	32.2
factored MIRA CLE	90.2	32.2	84.4	32.3

Work Approach

On a sample product I have taken few reviews as a reference for converting them as Multi-Dimensional Review Table (MDRT). To get the reviews from e-commerce site Amazon we have used API integration method and that reviews will be generated in the form of JSON format.

For Example we have considered 5 sample reviews for a product Mobile phone and generated the following sample table for one review.

Table 2

	Quality Parameters		
	Good	Average	Bad
Processor	5	0	0
Camera	0	3	0
RAM	5	0	0
Sensors	0	0	1
Battery	0	3	0

CONCLUSIONS

After generating all the key terms or indexed terms in the form of superior and inferior words we are going to generate a resultant table which shows the efficiency of the products in e-commerce websites. After table generation we are going to project the qualitative ratings of a product in the form of statistical representation.

REFERENCES

1. D.R.Kumar Raja, Dr S.Pushpa, B.S.Naveen Kumar "Multidimensional Distributed Opinion Extraction for Sentiment analysis – A novel Approach", *International Conference on Advances inElectrical,Electronics,Information,Communication and Bio-Informatics (AEEICB16)* 978-1-4673-9745-2 ©2016 IEEE.
2. Dependency Parsing Tutorial at COLING-ACL, Sydney 2006 Joakim Nivre Sandra K"ubler Uppsala University and V"axj"o University, Sweden. Eberhard-Karls Universit" at T"ubingen, Germany.
3. Sara Thomas Rosen University of Kansas,The Syntactic Representation of Linguistic Events Article <https://www.researchgate.net/publication/248029378> · January 1999.
4. Anusaaraka is an English-Hindi language accessing software <https://anusaaraka.iiit.ac.in/drupal/>
5. Jie Yang and Brian Yecies H Yang Mining "Chinese social media UGC: a big-data framework for analyzing Douban movie reviews" *Journal of Big Data* (2016) 3:3 DOI 10.1186/s40537-015-0037-9
6. "Sentiment Classification for Turkish Twitter Feeds using LDA" 978-1-5090-1679-2/16/\$31.00 c 2016 IEEE
7. D.R. Kumar Raja, Dr. S. Pushpa "A Survey on Privacy Preserving Data Mining Techniques" *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 10, Number 17 (2015)
8. Jeevanandam Jotheeswaran, Dr. S. Koteeswaran "Sentiment Analysis: A Survey of Current Research and Techniques", *International Journal of Innovative Research in Computer and Communication Engineering* Vol. 3, Issue 5, May 2015.
9. Shenghua Liu, Xueqi Cheng, Fuxin Li, and Fangtao Li "TASC:Topic-Adaptive Sentiment Classification on Dynamic Tweets", *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 27, NO. 6, JUNE 2015
10. Erfan Najmi· Khayyam Hashmi· Zaki Malik· Abdelmounaam Rezgui· Habib Ullah Khan "CAPRA: a comprehensive approach to product ranking using customer reviews" Received: 2 April 2014 / Accepted: 19 January 2015©Springer-VerlagWien2015